Reasoning with Reasoning

Using Faceted Browsers to Find Meaning in Linked Data

Steffen Hennicke[1], Kristin Dill[2], Gerold Tschumpel[1], Stefan Gradmann[3], Christian Morbidoni[4], Alois Pichler[5], Klaus Thoden[6]

(1) Humboldt-Universität zu Berlin, Unter den Linden 6, 10099 Berlin, Deutschland

(2) Österreichische Nationalbibliothek, Josefsplatz 1, 1015 Wien, Österreich

(3) Literary Studies Research Unit, Leuven, Blijde-Inkomststraat 21, 3000 Leuven, België

(4) Semedia, Università Politecnica della Marche, Piazza Roma, 22, 60121 Ancona, Italia

(5) Universitetet i Bergen, Sydnesplassen 12/13, Pb 7805, 5020 Bergen, Norge

(6) Max-Planck-Institut für Wissenschaftsgeschichte, Boltzmannstraße 22, 14195 Berlin, Deutschland

Table of Contents

1. Introduction

The use and application of digital research environments is of growing importance in the humanities. Within the discipline that has emerged out of the joining of the two fields of humanities and computing, the Digital Humanities (cf. Svensson 2009), there is an ever growing number of projects embracing Semantic Web technologies and Linked Data especially. As with all Digital Humanities endeavours, the question arises as to what extent the technologies developed in the context of computer science translate to the actual requirements of scholars in individual humanities disciplines (cf. McCarty 2005: 141). In 2009, Zöllner-Weber discussed the specific topics of logic reasoning and ontologies for the humanities. In her study regarding the use of inference tools in the domain of literature studies, she came to the conclusion that there are limitations to the application of such tools for humanities scholars. Not only does the use of these tools often require an in-depth understanding of mathematical logics, but the traditional scholarly activities in the humanities often involve "vague, ambiguous, or even contradictory" (Zöllner-Weber 2009: 10) information. In this sense, McCarty argued in 2005 that the benefits of computer science, which "focuses on combinatorics, syntax, and algorithms" and whose "guiding question is '*what can be automated?*'" fail to "address the humanities intellectually." (McCarty 2005: 141) This leads us to the question dealt with in this paper: What kinds of "reasoning" can humanists in fact apply with benefit to digital data, in particular, to Linked (Open) Data?

When we talk about "reasoning" in the context of the Semantic Web and the Digital Humanities we have to consider two principal senses of the term: the algorithmical use of the term as machine-supported inference of new knowledge, i.e. the creation of new relations in the graph, from a given knowledge base and the use of the term as the way

humans in general, and humanists[1] in particular, apply their styles of reasoning to the data and which inferences they draw (cf. Blanke et. al. 2013).

The first sense appears to be the most prominent interpretation and topic in Semantic Web digital humanities research[2] and often seems to obscure the second one. Semantic Web reasoning understood as large-scale machine-based inference, however, is not always accessible, feasible or even appropriate for applications and research questions within the digital humanities. All too often, the algorithmic potential of computers blocks the view of the seemingly simple but functionally useful Semantic Web tools available to the scholars already at present (cf. McCarty 2005).

The focus of research needs to be more inclusive with regard to the second sense and to examine if and how Semantic Web tools can support practices of reasoning and thinking about research topics typical for the humanities. This could necessarily build the basis for the application of reasoners down the road, but should also, and in the case of this paper primarily, serve to elucidate how scholars can work with such existing tools in the short term. According to McCarty (2005), "in the world of computational things we tend to value intricate, complex, algorithmically sophisticated tools, and so to undervalue what we actually have (...), [the] crude but functional" applications which allow us to explore new potentials (McCarty 2005: 112-113). These new potentials are arguably found in the data (cf. Oldman et al. n.d., or Gradmann 2013b) but also in the application of reasoning and problem-solving abilities of the human mind. In this sense, Deegan/Sutherland (2009) said the following in the preface to 'Transferred Illusions. Digital Technology and the Forms of Print':

---

[1] We use this term as a translation of the German word for "Geisteswissenschaftler" and not in the political sense.

[2] The topics dealt with in the International Summer School's yearly edition of "Reasoning Web" (z.B.2005-2012) are a testament to the prevalence of this notion of reasoning within the digital humanities.

> "As books do, computers measure, store and represent information; and as with books, it is still authors and readers who process this information as knowledge, making sense or nonsense of what the tool communicates. So far, computers cannot replicate common sense or creative thinking. The intelligence involved in processing information into knowledge is only a human skill."

'Human reasoning' can therefore be seen as an alternative to computer reasoning and a prerequisite for what McCarty calls 'human computing': "Human computing is computing that only proceeds with the essential and continuous engagement of human cognitive processes." (McCarty 2005: 147)

In this paper we, from a humanist point of view, will look at the application of human reasoning assisted by relatively simple digital tools, in particular tools to collaboratively and intellectually create and query Linked Data. In this context, we will specifically focus on the tools for the collaborative semantic annotation of digital resources that have been developed by Net7.[3] These include Pundit[4] (Grassi et al. 2013) and its family of applications: Korbo[5], Ask[6] and its built-in faceted browser for querying the semantic annotations made with Pundit. We will use these tools, in particular faceted browsers, as the basis for experiments being conducted with humanities scholars at two DM2E partners: the Wittgenstein Archives at the University of Bergen (WAB), [7] and the Georg-Eckert-Institute for International Textbook Research (GEI).[8] These experiments should shed light on how this tool may support, facilitate, or even hinder humanist reasoning in

[3] Net7 (http://www.netseven.it/en/) is the leader of the third work package in the EU funded project "Digitised Manuscripts to Europeana" (DM2E) (http://dm2e.eu/).

[4] https://thepund.it/

[5] http://www.korbo.org/

[6] http://ask.thepund.it/

[7] http://wab.uib.no/, with an open access edition of a part of Wittenstein's Nachlass hosted at http://wittgensteinsource.org/ and host of the Open Access Wittgenstein datasets www.wittgensteinsource.org (primary sources) and http://www.wittgensteinrepository.org/ (secondary sources).

[8] http://www.gei.de/en/home.html

a digital research environment based on Linked Data. Furthermore, the term "interactive reasoning" may characterise the practices that arises at the intersection between humanist reasoning and the Semantic Web reasoning by stressing the active involvement of the researcher during reasoning processes, i.e. how humanist researchers use Linked Data, or any data in a digital setting for that matter, to come to conclusions and find meaning with regards to their research questions. In our specific case, we will be focusing on a particular example of such interactive reasoning, namely faceted browsing. The aim of these experiments is not to achieve a systematic overview of all types of humanist reasoning that can be associated with Linked Data tools, but to investigate the way in which particular researchers may use their own styles of reasoning, typical of the humanities, to engage with Linked Data utilising simple exploration tools such as faceted browsers. We thereby strive to contribute a different perspective on the Semantic Web reasoning discourse.

First, we will introduce the context of reasoning within the Semantic Web domain. Then, we will discuss the term 'humanist reasoning' using the work of Holyoak and Morrison (2012), McCarty (2006) and Hacking (1985) as a basis. Afterwards, we will explore humanist reasoning use cases with DM2E partners at WAB, and GEI who are working with the DM2E tools Pundit and Ask. Finally, based on the observations that arise from these use cases, we will discuss potential Semantic Web reasoning applications in the first, computer-aided sense.

2. Reasoning

In this section, we will first discuss what the term "reasoning" traditionally means in the context of the Semantic Web and why its implementation in the humanities is difficult. We will then discuss what "reasoning" for scholarship on the basis of Linked Data in the humanities entails and explain our concept of "interactive reasoning" as a practical and

complementary alternative to the concept of automated “reasoning” in the Semantic Web.

2.1 Semantic Web “Reasoning”

In the Semantic Web, the term “reasoning” generally describes the ability for machines, so-called “reasoners”,[9] to automatically draw inferences from certain types of prepared data using formal logic and Description Logic; work in this area is related to the field of knowledge engineering (cf. Ludwig 2010). For this purpose, data is formalised in a triple structure based on the RDF data model. The semantics of that data is described by classes and properties which are formalised in ontologies. One simple way to describe the purpose of reasoning is that it is for “discovering new relationships”[10] in the existing data, this will be of importance when discussing how humans can interact with Linked Data in later sections of this paper. Here, we will discuss in more detail the basic elements required for Semantic Web Reasoning, which also ultimately play a role in understanding the “interactive reasoning” with Linked Data proposed here. These elements include RDF triples, RDFS, OWL, SPARQL, vocabularies and ontologies.

2.1.1. Linked Data Concepts

Semantic annotations according to the Linked Data paradigm at the most basic level consist of RDF-triples, which are simple statements about (Web) resources using an abstract syntax that is human and machine readable. This simple structure is analogous to basic sentence formation in natural language (especially English) and consists of a subject, a predicate and an object, where, according to the World Wide Web Consortium’s (W3C’s)[11] RDF-Primer, “[t]he subject and the object represent the two

[9] Cf. section 2.1.2.

[10] http://www.w3.org/standards/semanticweb/inference

[11] http://www.w3.org/

resources being related and the predicate represents the nature of their relationship".[12] Web resources as well as the relations connecting them are named with a Uniform Resource Identifier (URI), commonly in the form of an HTTP Uniform Resource Locator (URL) so that they can be unambiguously identified, easily found on the Web, and reused by other scholars.[13]

The following example shows the elements needed to create an RDF triple stating that a certain 'resource on the Web' (subject), here from a text published in Wittgenstein's Nachlass on Wittgenstein Source, 'discusses' (predicate/relation) the philosopher 'Plato'

(object). Plato is uniquely identified in this example using the URL from the corresponding DBPedia.org page.[14] The relation is uniquely identified with a persistent locator that has been catalogued and registered with purl.org.[15]

Figure 1: Basic triple structure: Subject - Predicate - Object.

Triples create a graph structure that can be infinitely extended by connecting nodes using new relations, making, for example, an object of one triple the subject of a new triple. The graph is often visualised as follows:

[12] http://www.w3.org/TR/2014/NOTE-rdf11-primer-20140225/

[13] http://www.w3.org/DesignIssues/LinkedData.html

[14] http://www.dbpedia.org/

[15] https://purl.org/docs/index.html, for example http://purl.org/wittgensteinsource/ont/discusses.

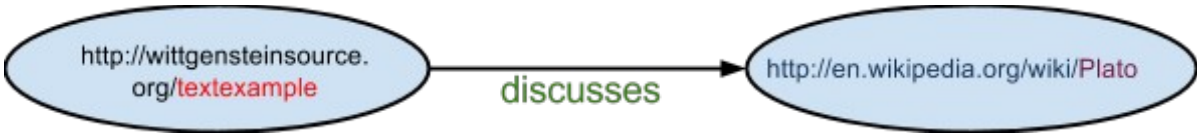

Figure 2: RDF triple as a graph.

The power of the graph can perhaps best be demonstrated by the ubiquitous Linking Open Data cloud diagram, which "shows datasets that have been published in Linked Data format,[16] by contributors to the Linking Open Data community project[17] and other individuals and organisations."[18]

[16] http://linkeddata.org/

[17] http://www.w3.org/wiki/SweoIG/TaskForces/CommunityProjects/LinkingOpenData

[18] http://lod-cloud.net/. Note that the DM2E data is also part of the LOD cloud.

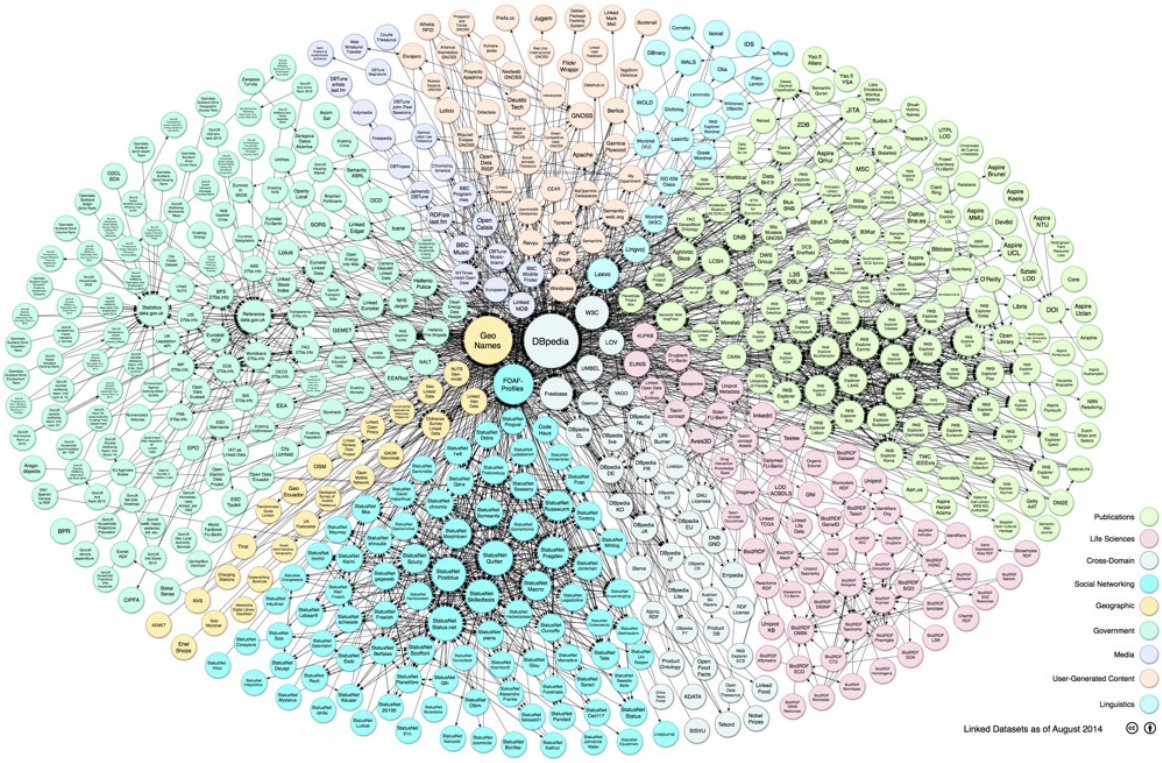

Figure 3: The LOD diagram from 2014.

The RDF Data Model is an abstract syntax[19]. In order for it to be useful for machines and humans in the modelling of data, it not only needs to be formalised as a concrete syntax (the exact rules for writing and storing the triples), but there also needs to be some consensus about the meaning of the predicates used and how they represent the

[19] cf. http://www.w3.org/TR/2014/NOTE-rdf11-primer-20140225/.

relationship between the subjects and objects. For the former, RDF vocabulary and RDF syntax languages such as Turtle[20] or RDF/XML[21] are used; they will not be considered in detail in this paper. For the latter, schemas, vocabularies and “ontologies” have been established.

The Resource Description Framework Schema (RDFS) extends RDF by providing “mechanisms for describing groups of related resources and the relationships between these resources.”[22] More specifically, RDFS provides a vocabulary for defining classes and properties, and to create subclass and subproperty taxonomies. Furthermore, the domain and range of properties can be specified. These constructs have simple predefined semantics which already allow simple kinds of reasoning such as transitive reasoning along subclass relations. RDFS therefore provides a basic “grammar” (Gradmann 2013) for the semantic modelling of data, it cannot, however, cover more complex modelling needs. For this purpose, there is the Web Ontology Language (OWL),[23] which is now in its second version as OWL2.[24] For the purposes of this paper, we will use the term OWL to refer to both editions.

OWL and its sublanguages (OWL Lite, OWL DL and OWL Full) facilitate the development of ontologies “by providing additional vocabulary along with a formal semantics” based on Description Logic (DL) for this purpose.[25] Ontologies created on the basis of OWL not only help to structure the knowledge in a certain domain, they also allow for the inclusion of more logical constructs which can then be ‘understood’ and processed by machines. They therefore serve as the basis for machines to complete

[20] http://www.w3.org/TR/2011/WD-turtle-20110809/.

[21] http://www.w3.org/TR/REC-rdf-syntax/.

[22] http://www.w3.org/TR/rdf-schema/

[23] http://www.w3.org/TR/owl-ref/

[24] http://www.w3.org/TR/owl2-overview/

[25] http://www.w3.org/TR/owl-features/

reasoning tasks. Gruber (1993) describes the term ontology as “an explicit specification of a conceptualization.” In other words, creating an ontology is a way to represent and contextualise a certain section of reality (cf. Gradmann 2013: 222).[26] One could also say that the ontology helps to create a knowledge base that “store[s] knowledge about the real world” in a certain domain (cf. Ludwig 2010).

One common way to query the data stored as triples is by using the W3C recommended querying language SPARQL,[27] an acronym of “SPARQL Protocol And RDF Query Language”, in connection with a SPARQL endpoint interface. Accessing the data in this manner demands not only a knowledge of the language itself, but also previous knowledge of the types of entities and relations in the triple store as well, which can be quite a barrier for those not familiar with the dataset when trying to query the data.

2.1.2. Reasoning and Inference using Machines

One major area of interest in the Semantic Web domain is being able to process the formalised “representation[s] of terms and their interrelationships”[28] expressed using ontologies by applications so that logical inferences about the data can be made on a large scale. This is, as mentioned above, generally considered “reasoning” in the Semantic Web sense and is geared towards automatically finding implicit knowledge in the data. Reasoners for automatically finding such logical connections inherent in the data created have to be tailored to the particular ontologies and needs of the specific research question and domain (cf. Gardiner 2006). In their “Comparisons of Reasoners for large Ontologies”, Dentler et al. (2011) provide a solid definition of the term “Reasoner”: “A reasoner is a program that infers logical consequences from a set of

[26] For example http://www.w3.org/TR/owl2-primer/#Class_Disjointness

[27] http://www.w3.org/TR/rdf-sparql-query/

[28] http://www.w3.org/TR/owl-features/

explicitly asserted facts or axioms and typically provides automated support for reasoning tasks such as classification, debugging and querying."

As mentioned above, a reasoner uses logic based in mathematical theory to infer new information automatically from the existing triples in the graph. The W3C provides a simple example of how this type of inference works; note its similarity to the philosophical syllogism:

> "The data set to be considered may include the relationship (Flipper isA Dolphin). An ontology may declare that 'every Dolphin is also a Mammal'. That means that a Semantic Web program understanding the notion of 'X is also Y' can add the statement (Flipper isA Mammal) to the set of relationships, although that was not part of the original data. One can also say that the new relationship was 'discovered'. Another example is to express that fact that 'if two persons have the same name, home page, and email address, then they are identical'. In this case, the 'identity' of two resources can be discovered via inferencing."[29]

This type of reasoning relies primarily on the consistency of the outward form of the statements in the dataset, which not only have to be apophantic, but also have to accurately represent the objective reality of a certain domain on some level to provide the conditions for relevant conclusions (cf. Zoglauer 2008). The original statements also have to have been at some time provided by humans based on their prior knowledge and reasoning about the domain. A Semantic Web reasoner cannot, for example, interpret whether the underlying information (presuppositions) that has been represented in the rule set or vocabulary is actually factual, valuable, sensible, objective, unbiased, relevant or even useful. Only a human has the ability to acquire knowledge and determine meaning. Therefore, a dataset could state, for example, that 'Flipper isA Dolphin' and the ontology that 'Every Dolphin is also a Bird'. The resulting information 'Flipper isA Bird' deduced from the initial premises would be logically coherent (cf. Zoglauer 2008: 9), but for the human observer of course mere nonsense.

[29] http://www.w3.org/standards/semanticweb/inference

Humanists, however, are not necessarily interested in the form, but in the meaning of statements (cf. Oldman et al. n.d.) about objects in their domains.

This type of Semantic Web reasoning can be used, according to the W3C,[30] for "improving the quality of data integration on the Web", or may help the researcher to "automatically analyse the content of the data" or to "discover possible inconsistencies in the (integrated) data". In certain circumstances it may even be used for "discovering new relationships". This process, however, is contingent on facts being explicitly stated by the scholars and is therefore limited to their ability and willingness to do this, the socio-historical context in which they do this, and the quality of the information contained in the modelled data. Such uses for Semantic Web reasoning are therefore, in the short term, not necessarily as promising for humanist research using Linked Data, not in the least because this type of reasoning is limited in scope.

2.1.3. Difficulties of Semantic Web Reasoning in the Humanities

There are several reasons why this type of machine-supported inference has limited use and relevance from the perspective of a humanities scholar. We will mention three obvious ones here.

First, understanding and utilizing this type of machine-aided reasoning requires at the very least a basic knowledge of concepts and skills that are uncommon in most humanities domains. They include, but are not limited to computer programming and querying languages, Linked Data concepts such as the ones previously introduced, database management, ontology creation and knowledge representation, the use and implementation of inference machines, and formal logic in its mathematical expressions such as Description Logic. Even power users including the Digital Humanist might have a steep learning curve for some of these skills. Persuading a humanist to take the time

[30] http://www.w3.org/standards/semanticweb/inference.

to learn how to apply Semantic Web reasoning requires at the same time a clear understanding of the benefits that will be reaped for her domain and specific research interest.

Second, the objects of study, the types of research questions, and the methods found in the humanities are not always compatible with the Semantic Web reasoning paradigm. As mentioned above, mathematical logic is mainly concerned with the form of statements while, in contrast, humanists have complex and often contradictory research objects (cf. Oldmann et al. n.d.) and are interested in layers of meaning. For example, a historian has little use for creating labour intensive knowledge representations that allow a computer to “infer” that Flipper – a fictional character from a TV-Show – is a mammal. Unambiguously defining an object is rather the concern of the applications of natural sciences, where such inference machines have been successfully implemented. The historian is however perhaps more concerned with what this fictional character might represent to the audience of one or more time periods. This requires extensive knowledge about several domains such as the culture of the society in general and television in particular, the history and culture of the reception of the show, the particular language of the imagery used in the series and its relationship to other shows. The information (datasets and ontologies) required for an algorithm to automatically come to the historian’s conclusion would be difficult to create and implement. A conclusion the historian might draw such as “Flipper isA(n) aquatic Lassie” is metaphorical, highly subjective, and neither true nor false, making it not a necessarily good candidate for the premise or potential conclusion of a formal logical statement. At the same time, this does not mean that this assertion based on analogy is necessarily unfounded or irrelevant. As Pesce (1999) states “meaning cannot be counted”, i.e. translated into an unambiguous language that the computer can process.

Third, the demands placed on this type of reasoning are ambitious – "to solve problems in domains that ordinarily require human logical reasoning" (cf. Ludwig 2010) – but the machines' ability to facilitate such inference on a large scale (and perhaps also to a humanist's standards) is often limited and contentious (cf. Zöllner-Weber 2009, cf. McCarty 2005). In this context, the modelling of a domain in such a way so that inference machines can eventually create valuable knowledge from it is an activity that is dependent on a large investment of human reasoning in the first place. The information to be "discovered" has to be preconfigured in the knowledge representation. Perhaps the combining of the dataset with the ontology can lead to the computer being able to "infer" that "Flipper isA Mammal", but only because humans 'know' this in the first place. What does a humanist gain by intensively modelling a domain so that the computer can discover what she already knows?

2.2 Reasoning in the Humanities

Providing an extensive analysis of humanist reasoning practices would go far beyond the scope of the paper and would be an elaborate scholarly endeavor in and of itself. Our purpose here is instead to point to other ways of thinking about the term "reasoning" for the Semantic Web that are more familiar to the humanist, so as to shift the weight of the discussion towards a position that includes the value of humanist input about thinking about data created in the Web environment, especially in environments using Linked Data. For this reason we will first provide a much broader definition of reasoning than the one given above. This will hopefully help to highlight the potential of humanist ways of looking at the data, which we will then explore in our use cases in the next section.

The Oxford Handbook of Thinking and Reasoning generically defines the term "reasoning" as follows: "Reasoning, which has a long tradition that springs from philosophy and logic, places emphasis on the process of drawing inferences

(conclusions) from some initial information (premises)" (Holyoak and Morrison 2012: 2). According to the authors, reasoning is intrinsically related to but not necessarily synonymous with the act of thinking in general and closely tied to many other mental activities such as judgment, decision making, creative thinking and problem solving (cf. Holyoak and Morrison 2012: 2). Although the rigorous confines of formal logic are used in some reasoning practices such as those of (analytic) philosophy and mathematics, there are other practices or 'styles' of reasoning that place much less emphasis on this (cf. McCarty 2006, cf. Crombie 1994 and Hacking 1985). Indeed, logic is sometimes considered an attempt to provide a normative model for the reasoning process (Holyoak and Morrison 2012: 4-5) or a "grammar of thought" (Zoglauer 2008: 9), but it is not to be confused with all of the complex cognitive processes and scientific practices involved in reasoning itself.

Crombie (1994: 155) explicitly makes a distinction between logic and reasoning. He states: "First, I observe that by reasoning I don't mean logic. I mean the very opposite, for logic is the preservation of truth, while a style of reasoning is what brings in the possibility of truth or falsehood." While the exact wording of this distinction may be contentious, we find the general tenor of the statement to be of value, that logic and reasoning are not synonymous: logic is a tool that can be used to make sure that arguments are sound, but reasoning involves the entire process of coming to conclusions and is dependent on different scientific cultures.[31] McCarty (2008: 12), referencing Crombie (1994), talks about different cultural practices which have evolved

[31] To be sure, there is much overlap between styles, modes or practices of reasoning and it should not be implied here that these types of reasoning are somehow less rigorous than the types of reasoning specific to mathematics, philosophy and the natural sciences (cf. Holyoak and Morrison 2012: 11). As Holyoak and Morrison (2012: 3) point out, Thomas Hobbes, in designating reasoning as form of "reckoning", equated it with "'computations,' as in arithmetic calculations". This notion of thinking and reasoning has persisted within the sciences (Holyoak Morrison 2012: 3ff.) and is not unfounded. Indeed this implies a type of computation that is at the heart of every reasoning practice.

to help humans come to conclusions as ‘styles’ of reasoning. In this respect he provides us with the following list:

> “The simple method of postulation exemplified by the Greek mathematical sciences; The deployment of experiment both to control postulation and to explore by observation and measurement; Hypothetical construction of analogical models; Ordering of variety by comparison and taxonomy; Statistical analysis of regularities of populations, and the calculus of probabilities; The historical derivation of genetic development.”

Keeping this in mind, Holyoak and Morrison (2012) discuss a number of different aspects that need to be considered when talking about “reasoning” in a wider sense. These include not only the different scientific approaches that have normalised reasoning processes, but also the different methods of coming to conclusions (inductive, deductive, abductive), the intricacies of judgement and decision making, the impacts of language and culture on reasoning, and different modes and practices of thinking.

This brings us to our plea for widening the understanding of reasoning in the digital humanities, especially related to the Semantic Web, to include other styles of reasoning with the data than the purely computer-oriented ones mentioned in the previous section. Since human reasoning is ultimately at the basis of any reasoning programme, understanding what reasoning practices humanists may engage in with the available data can also ultimately help the future implementation of reasoning in the computer science sense. It is also important to discover the ways humanists are able to reason (come to conclusions) using the simple tools and functionalities immediately available to them.

For our purposes then, the definition of Semantic Web “reasoning” will be extended to include any process of interaction with (Linked) Data and the resulting graph that leads to the discovery of new information and the potential creation of new triples. Note that

this definition does not restrict reasoning to drawing inferences, but still has a focus on coming to conclusions. This interaction can be either driven by human or computer interaction. In our case we will concentrate on human interaction with the graph, specifically on the technology of the faceted browser. This will be discussed in more detail in the following section.

2.3 Interactive Reasoning

Although we have made the definition of Semantic Web reasoning for this paper very broad, our vision for exploring what we call “interactive reasoning” for the Semantic Web is more narrow in scope. Here we will concentrate on how humans can use the simple but effective tools that exist for exploring and exploiting the Linked Data graph in order to come to new conclusions (i.e. create new triples) about the data. Through our use cases we have discovered the potential for studying “interactive reasoning” for scholarship on the basis of Linked Data in the humanities in this regard on at least three levels. Reasoning as a human cognitive activity is involved in

1. selecting, modifying, or creating (annotation) vocabularies for particular data sets and research interests;

2. applying the (annotation) vocabulary by annotating resources;

3. exploring and assessing the data by visualising and querying the graph that has been established through the creation of an (annotation) vocabulary and the annotation of a data set.

This third step can not only be carried out by a reasoner, but also by a human using other tools such as faceted browsers. Our case studies will reflect these areas of humanist “interactive reasoning”. The next section will discuss reasoning with faceted browsers in general.

Using, among other things, the scholarly platform based on the Pundit family of tools, DM2E has been conducting research into current and potential scholarly practices with Linked Data (cf. chapter 2). Central to the current task on "reasoning" has been to further explore the useful but still limited capabilities of the Ask faceted browser by having data created by DM2E scholars made searchable using additional faceted browsers implemented for their specific purposes. Scholars could then use the results to make inferences about the data created, potentially coming to new conclusions, discovering new information and creating new links.

An overview of the functionalities of the Pundit family of tools used for creating the datasets in the case studies can be found in other DM2E publications.[32] Here, we will briefly discuss Ask, as it includes the potential of the faceted browser. Ask provides a domain independent view on annotations (Cf. D3.3) created in Pundit. It is used for managing personal notebooks containing annotations, viewing notebook contents, and providing a basis for simple vertical visualisations. The notebooks faceted browser, which allows for any number of notebooks to be searched dynamically, is a very powerful feature for analysing a corpus of annotations. But since Ask is domain independent, it provides only a generic way of exploring the whole graph and the facets are limited to the particular instance data of the subjects and objects, the properties, and class types.

A faceted browser can however, be tailored to a specific dataset. In general, it is an application that allows the user to access data using different filters. The data can then be combined and recombined in different ways depending on the chosen filters, providing novel ways of looking at the data. Faceted browsers have a distinct advantage for scholars from humanities' domains wanting to explore and query information stored as Linked Data. For one, they provide an immediately accessible but structured

[32] Cf., for example, Grassi et al. (2013).

visualisation of the specific subjects, objects and predicates in the data set. Browsing, in contrast, an RDF/XML document per se is not the easiest way to make sense of the data. In addition, if a faceted browser is provided, other scholars who are perhaps less technically inclined can obtain this overview of the contents and structure of the data set without needing any knowledge of programming languages or query languages such as SPARQL. They can therefore solely concentrate on comparing their understanding of the domain with the one represented by the data set. In short, a faceted browser facilitates reasoning by permitting the scholar to relatively quickly identify, in a given dataset, the data and metadata most relevant for the pursue of a specific research question by iterative processes of selecting and deselecting given facets. The facets offered at each time for selection will always be a result from exactly the previous selections and thus guide the scholar through and document the research path. Such reasoning by faceted browsing is to us an example of a kind of reasoning that humanists can perform with benefit in the context of the Semantic Web and the Digital Humanities, even though it is not fully automated reasoning, and thus not Semantic Web reasoning in the standard sense.

For our experiments, we had scholars load their data into faceted browsers tailored to their specific research data and domain. In the next section we will discuss reasoning with Linked Data based on two case studies, including ones utilising faceted browsers.

3. Reasoning Use Cases

For our research into “interactive reasoning” with Linked Data we analysed case studies (Wittgenstein Archives at the University of Bergen, Georg Eckert Institute) from two separate digital humanities domains (philosophy, history), for which we provided self-documentation forms and conducted expert interviews. This section will discuss the two case studies in detail and then explore the reasoning scenarios they entailed.

3.1 GEI Case Study

For the first case study to be examined in this paper, DM2E worked with a scholar at the Georg-Eckert-Institute (GEI), a DM2E associate partner, from the field of educational history. An experiment with Pundit and its components was set up for this purpose. The experiment took place between September and November 2014. The scholar was given an introduction to important Linked Data concepts and methods and asked to define a relevant use case from his particular interest area in the field of history. The participant was asked to determine the nature of the semantic annotations to be established and the sources to be used based on his usual research methods. Then, the participant created about 250 semantic annotations on relevant historical digital materials, which resulted in small but meaningful graphs. The process can be described as a basic and intuitive translation of common research methods from the humanist field of history into the Linked Data paradigm for the purpose of answering research questions specific to a humanist scholar's area of interest and expertise. For this reason it served, among other things, as an use case for analysing the way in which humanists might want to "reason" with Linked Data. For the purposes of the "reasoning" aspect of the experiment, the scholar was asked to explore the graph using a faceted browser and given a self-documentation form to complete (Addendum), which had been previously prepared by the members of DM2E in the context of Task 3.4.

3.1.1. Method

The source documents for the experiments were taken from the digital library of GEI and consisted of different types (e.g., protestant or catholic) of historical school books from Germany published from ca. 1850 to 1900.[33] The (RDF) metadata of these sources

[33] The corpus of GEI-Digital was created by the Georg-Eckert-Institute for International Textbook Research Member of the Leibniz Association. It was created to be used by different scientific communities (e.g. historians, education researchers). The aim is to permit an easy access to this kind of source

have also been provided to Europeana via DM2E. The particular research method was hermeneutic and involved closely reading the historical sources and identifying various topoi (salient terms), and their connotation (positive or negative) and presentation in the different school books. Specifically, the experiment focused on the questions: “Which topoi appear in which textbooks?”, “How are they connotated?” and “In which context have they been set?” Topoi were annotated according to several criteria including the nature of the connotation. The desired result of the annotations was to be able to compare the topoi in different texts over time, assessing, for example, which topoi can be found in which documents and how the connotation and frequency of certain topoi change over time.

The scholar is part of a research group that works on and with textbooks and juvenile literature of the nineteenth century. Because textbooks are semi-official documents that were read by wider parts of the Germans during their formative years, his group tries to find the representations of the world and the nation and the description of historical processes that were offered by the state to its future citizens. So, they search for representations of the nation and the globalised world. Also, they look for representations of change, crisis, religious conflict, social change and similar events.

The experiment involved a three-step process consisting of a 1) source critique (documented with Pundit), 2) content analysis (documented with Pundit) and 3) exploration of the data with a faceted browser (Ask). The first two steps are based in the hermeneutical method and translated to the Linked Data paradigm. A detailed analysis of the results will be available in the form of a paper in the future. The last step involves

---

material; to capture them in form and content and to make full-text versions available to a wider, international user group. The parallel aim is long-term sustainability of the books themselves. It contains images, bibliographical data, context, full text, a pdf version, esp, mets/mods

using the automatic visualisation and combination functionalities of the faceted browser to evaluate the results and therefore the reasoning process involved.

The source critique consisted of skim reading the texts and establishing an initial contextualisation. This involved selecting sources from the digital library of GEI, reading the sources based on experience and hypotheses, identifying noticeable entities in the text and collecting potentially interesting texts in a notebook as a sample by creating annotations with Pundit. The annotations used in the first stage of the source critique identified facts such as

- place of publication
- year of publication
- edition
- author
- publisher
- religious attribution of the text (e.g. protestant, catholic, neutral)
- school type (e.g. girl's school, boy's school, gymnasium, teacher's seminar)
- regional attribution (e.g. for schools in Baden)
- discipline (e.g. history, geography.)

These basic attributes were the building blocks for establishing context by connecting them with further facts about the corresponding historical environment, which included

- relevant historical events, such as the foundation of the German Reich in 1871
- historical periods: ~1850-1870, 1871-1884, 1885-1900, 1901-1914, 1914-1918
- names of smaller sub-periods (German: "Querperioden"), such as the period of the Socialist Laws (German: "Sozialistengesetze") from 1878 to 1890.

In addition to annotating attributes of elements within the documents and contemporary events corresponding to the time period they were in use, the scholar was also interested in analysing the content according to the implicit values and opinions contained within them. For this purpose, the second step involved content analysis, which meant looking for evaluative concepts. First, pertinent “topoi" and "connotations" were identified and annotated using RDF triples with the form

- :x :is_a :topoi
- :x :is_used_as :connotation (e.g. “anti-secular”, “anti-religious”)
- :x :is_used_with :positive_connotation, negative_connotation

Second, paraphrases of the topoi were identified and annotated:

- :y is_a :paraphrase
- :y :refers_to :topoi

This process in terms of Linked Data can be described as the creation of a small-scale vocabulary for the purpose of documenting evaluative statements found in historical texts. The results of this experiment can be currently viewed in the faceted browser.[34] One example of the triples created can be illustrated with the following screenshot taken from the scholar’s Pundit Notebook entitled “Welt der Kinder”:

[34] http://demo-search.thepund.it/.

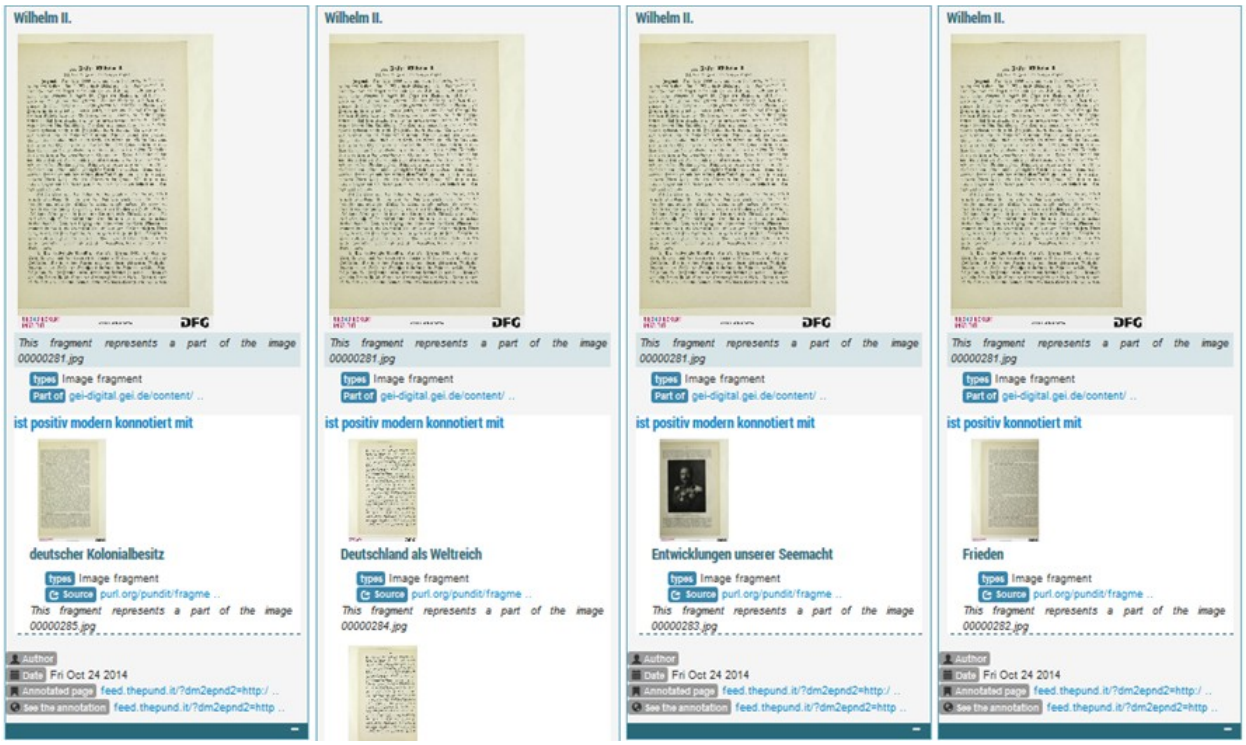

Figure 4: Triple-display, "Welt der Kinder" notebook in Pundit.

The four textual examples all come from the same textbook: "German History from the Migration Period to the Present" (German: "Deutsche Geschichte von der Völkerwanderung bis zur Gegenwart") by Ludwig Kahnmeyer, Adolf Dilcher and Hermann Schulze, which was published 1913. They deal with the "topos" of Wilhelm II, the German Emperor who ruled the German Empire and the Kingdom of Prussia from 1888 to 1918. The textbook therefore was published during his reign. The historian created several annotations for this topos (triple subject), which occurs on several different pages of the textbook (pp. 281-285). It is positively connotated (triple predicate/relation) with the concepts (triple objects) "development of our naval power" (German: "Entwicklungen unserer Seemacht"), "Germany as a world power" (German: "Deutschland als Weltreich"), "the possession of German colonies" (German: "deutscher Kolonialbesitz") and "peace" (German: "Frieden").

This small section of the work done in this experiment shows an example of the results of the researcher's methods. For one, we see several aspects of the official German state propaganda of the time, which here are expressed as pride in the military and an attribution of state policies to the figure of the Emperor. In the larger context of state propaganda in general, we can clearly see the often overlooked contradictory nature of ideology: a political figure can be positively connotated with militaristic concepts such as "naval power" or "colonialism" and, at the same time, with "peace".

It is evident from the example annotations concerning Wilhelm II given above that the first two steps of the experiments based on the historian's hermeneutic method involving the close reading of a source were able to deliver telling results about the research object. For these two steps, the scholar did not necessarily need the aid of any computer technology at all. However, the simultaneous utilisation of simple Linked Data tools (creating semantic annotations with Pundit) provided him with the basis for new ways of storing and displaying his data that can lead to novel ways of looking at, working with and reasoning on the results obtained.[35] This was explored in the third step of the experiment: exploring the data with Ask.

Although the scholar did not create an elaborate ontology to represent his particular domain, he did establish a small vocabulary to explicitly document both historical facts and statements about meaning concerning the content of his research object. The evidence behind his conclusions as well as the conclusions themselves, captured in the form of triple statements, could therefore be automatically recalled using Ask and a simple faceted browser, PunditSearch,[36] built specifically for this purpose.

---

[35] The scholar did not involve heavier Semantic Web Reasoning methods such as the creation of an ontology, computer algorithms or formal logic to come to his conclusions or achieve his results.

[36] http://demo-search.thepund.it/

In the third step, the scholar was asked to use the faceted browser to query the data with regards to a specific research question about the source and the Linked Data created in the first two steps. The faceted browser allows to incrementally filter triples based on the instances found in the subject, predicate, and object position of triples, and according to classes. The browser adjusts the display of triples matching the selected facets. Below is a screenshot of the PunditSearch faceted browser.

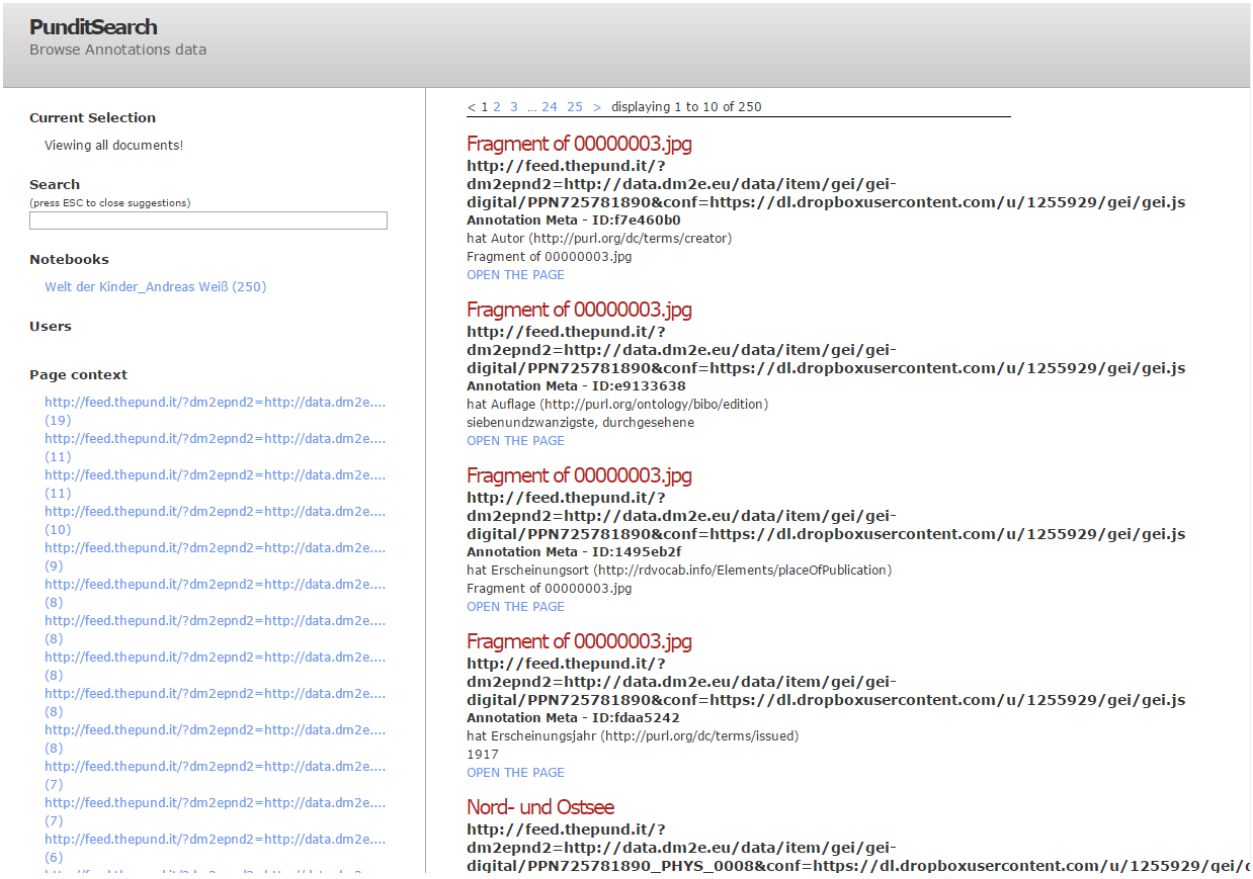

Figure 5: Screenshot of the PunditSearch faceted browser displaying the GEI dataset

At the same time, the scholar was given a self-documentation form in which he was asked to describe and reflect on the process. The following is a summary and analysis of the documentation.

Our scholar decided first to obtain an overview of the object labels, as he found oversight of the triples created to be difficult. After browsing the available object labels, he chose to take a closer look at the most numerous (with 5 instances), the “German Empire”. The list of related (triple) subjects returned by the browser surprised him, as the German Empire was only connoted with what he felt were “internal topics” (national as opposed to international). He had expected the object “German Empire”, however, to also be compared with “external” topics such as “France” and other “Empires”.

This first look at the object label group raised questions for our historian, which could be considered part of a reasoning process that can lead to the creation of new triples. As the scholar looks at the subjects returned as a result of the faceting browsing, he compares them to his expectations, which are based on real-word knowledge. As a consequence, he discovers a similarity in the results that occurs to him because of the absence of what he expected to be in the results. This is quite a dynamic, intuitive and partially serendipitous reasoning process. Before looking at the group, the scholar might not yet have known what his expectations were, i.e. that the category of “national” versus “international” is relevant when describing the way in which textbooks talked about the “German Empire”. He now can make this implicit information explicit by creating a new triple expressing this (a product of reasoning as defined in this paper). In addition, he already has a result and can use his knowledge to further determine what it means: “One way the topos of the “German Empire” is constructed/talked about in the school books is by listing positive attributes related to national subjects”.

The exploration (faceted browsing) of the object label “German Empire” returned only positively connotated results. This unexpected pattern led the scholar to take a look at all the topoi that are connotated negatively (43) and all that are connotated positively (63). This additional step occurred out of a desire to gather more data to understand the results. Here, an unexpected pattern causes the scholar to look for other significant

patterns, relationships and groupings in the dataset of positively and the negatively connotated topoi.

In the process he discovers that one of the negatively connotated subjects is France. Being a European country, this subject can be considered a similar but distinct counterpart to the object label “German Empire”. The scholar’s choice to look at France therefore has a basis in analogy. The scholar has a look at the topoi (subjects) associated with France, discovering that the negative connotation has to do with France’s naval power and its open borders. Combining this with his real-world knowledge, he comes to the conclusion that this contrast is an expression of “military and political rivalry.” He compares this new information to his previous results about the “German Empire” and also comes to a conclusion, that an antagonism is created in the textbooks between the “German Empire as a supporter of peace and France as an aggressive and potentially dangerous neighbour”.

Once again the results make the scholar want to explore more data. With each conclusion about a certain subject he extends the search to other analogous elements of the dataset in an iterative process. The scholar therefore resets the facets to search for what connotations have been made about other neighbours of the “German Empire”. The results lead the scholar to believe that most countries are constructed as potential rivals of Germany.

The scholar’s final conclusion from this search is that “the German nation is represented as a modern peaceful one that exists unfortunately in a dangerous environment. And crises and potentially dangerous changes loom everywhere!”

3.1.2. Results

In the scope of this paper it is impossible to make sweeping statements about what types of Semantic Web reasoning all humanists want to see employed with Linked

Data, but we can use this case study to make empirical observations about the practice or style of reasoning found at the intersection between traditional humanistic research practices and those (that will be) made possible through the use of Linked Data tools.

In order to be able to determine the style of reasoning expressed in this use case, three relevant aspects of the experiment should be discussed. First, one of the difficulties of Semantic Web reasoning mentioned previously in the paper was that the research objects and research questions of humanities scholars are qualitatively different than those of scholars of mathematics and life sciences. For this reason, the first aspect considered will be the scholar's chosen research object and research question. The second aspect will discuss the underlying research method used in the creation of the Linked Data, which can be described as being at the intersection of humanist and Linked Data methods. Lastly, we will discuss how our scholar used the faceted browser tool to come to conclusions about the RDF data set he created and how he assesses the method.

A look at the object of research and research question in this case study supports the idea that humanists are interested in meaning. The scholar was interested in historical facts, but more importantly he wanted to study the construction and expression of opinions, values, worldviews, and biases in the historical school books. Historical facts were important for providing the context of the value 'statements', but addressing their meaning to the authors and potential influence on the recipients of the works was the most important aspect. As a result, the small vocabulary created by the scholar was primarily tailored to the documentation of attitudes. Although statements created are useful, they are not necessarily axiomatic.

The underlying research method and therefore perhaps also reasoning style of the scholar was by self-admission hermeneutic. It involved close reading and textual

interpretation based on real-world knowledge, uncovering subtext, and searching for meaning. This was then combined with the Linked Data methods by having the scholar explicitly state the results of this close reading in his annotation. Meaning is therefore interpreted in this process and stated using triples.

The scholar's self-documentation of his interaction with the graph using the faceted browser gives us some insight into how humanists can use their methods to interact with and ultimately reason with Linked Data. More specifically, we are shown how they can draw conclusions from the data using the faceted browser. Of course this type of reasoning is limited to and contingent upon the specific technological implementation. It is, however, instructive in uncovering a complex and dynamic humanistic reasoning process using Linked Data.

In general, we have noticed a main procedure of reasoning with Linked Data using facets that is iteratively repeated: the scholar analyses the results (of applying certain facets to the data) by comparing them with his real-world/scientifically acquired knowledge in different ways and creating hypotheses about them based on this. With the faceted browser he can then re-shuffle facets to find how other combinations undermine or support hypotheses and to look for answers to new questions that arise. This reasoning procedure is conducted on the basis of statements the scholar created himself, i.e. the annotation vocabulary and instance data, which allows him to better comprehend the context of what he sees.

Real-world knowledge means keeping in mind that results shown in the faceted browser are reflective of certain assumptions and biases explicitly and implicitly addressed within the texts that can be determined by considering their context. Context is, of course, more than just referencing the name and vita of the author(s) and sponsor(s) or the dates of the time period it was written, but a deep understanding of what hidden

agendas the authors had, what values they were trying to perpetuate, and what this all means for people living at the time of publication as well as today. The significance or relevance of certain aspects of context shifts according to the question asked, who asks it, for what purpose and the results given.

When looking at the results, the scholar observes, for example,

- (relevant) patterns in the data
- a pattern/information that supports expectations
- a pattern/information that contradicts expectations made from judging the context
- unexpected information
- salient information
- anomalous information
- absurd information
- analogous resources/information
- antagonistic resources/information

and he compares and contrasts this with real-knowledge and experience.

The scholar's own criticism of the resulting reasoning process is valuable for assessing the value and potential uptake of this kind of "interactive reasoning" with Linked Data.

On a positive note, the scholar cited several positive aspects of this approach. He felt it aided thoroughness through forcing the repetition of statements. It aided his ability to quickly compare the data created from the close reading of several sources. A hypothesis could be instantly tested and results automatically reproduced. The recall of relevant information was therefore much faster.

3.2 Wittgenstein Ontology Case Study

The second use case provides a further perspective on the reasoning topic. In contrast to the use case with the GEI, the researchers at the Bergen Wittgenstein Archives (WAB) had previously created an ontology (using RDF-triples) for use with the data in its archive, the so-called Wittgenstein Ontology (WO). This ontology was developed by digital humanists with knowledge of Semantic Web technologies. The use case focused on the characteristics of the ontology and on the interaction of two scholars with the ontology using the faceted browser.

3.2.1. Method

WAB is a partner of DM2E, providing a digitised edition of Wittgenstein's Nachlass, which is produced from WAB's machine-readable version of Wittgenstein's Nachlass.[37] The Wittgenstein Nachlass amounts in total to ca. 20,000 pages, while the Wittgenstein Source corpus on wittgensteinsource.org includes a 5000 page selection from this larger corpus. Wittgenstein Source was created in the framework of the Discovery project[38] by WAB for Open Access Wittgenstein research. It contains English and German manuscripts and typescripts from Wittgenstein's Nachlass in facsimiles and as diplomatic and normalised text editions. It also contains metadata and short descriptions of these items.

The WO is linked with Wittgenstein Source, partly through Pundit. The Linked Data representation of the ontology was created by WAB for both internal and external use. Internal use includes checking of metadata comprehensiveness and consistency, external use (by researchers) includes searching and browsing of metadata. The ontology was intended primarily to assist Wittgenstein research. It includes classes for primary and secondary sources, concepts and persons. The lowest subclass of a

[37] Ref. to the Bergen Electronic Edition.

[38] http://wab.uib.no/wab_discovery.page and http://www.discovery-project.eu/home.html.

Wittgenstein primary source is the Bemerkung; the Bemerkung denotes, roughly speaking, a single Wittgensteinian remark (German: “Bemerkung”). Instances of the different classes are interlinked with each other via properties / predicates.[39]

For querying the data set of the WO, Net7 worked with WAB to create the Wittgenstein Ontology Explorer,[40] which is a semantic facets browser using the open source software Ajax-Solr.[41] Users can choose from the following facets (subjects, objects and predicates in the ontology) or search for terms in the facets using a search bar:

- Type (source category - primary or secondary)
- Published in (work)
- Part of (manuscript, typescript ...)
- Date (of remark)
- Source (secondary or primary)
- Refers to (person)
- Discusses (topic)
- Other version (of remark)

Below is a screenshot of the faceted browser, in which no facets value have been selected.

[39] Cf. Pichler and Zöllner-Weber (2012).

[40] http://141.20.126.236/dm2e/ajax-solr/examples/wab/.

[41] https://github.com/evolvingweb/ajax-solr.

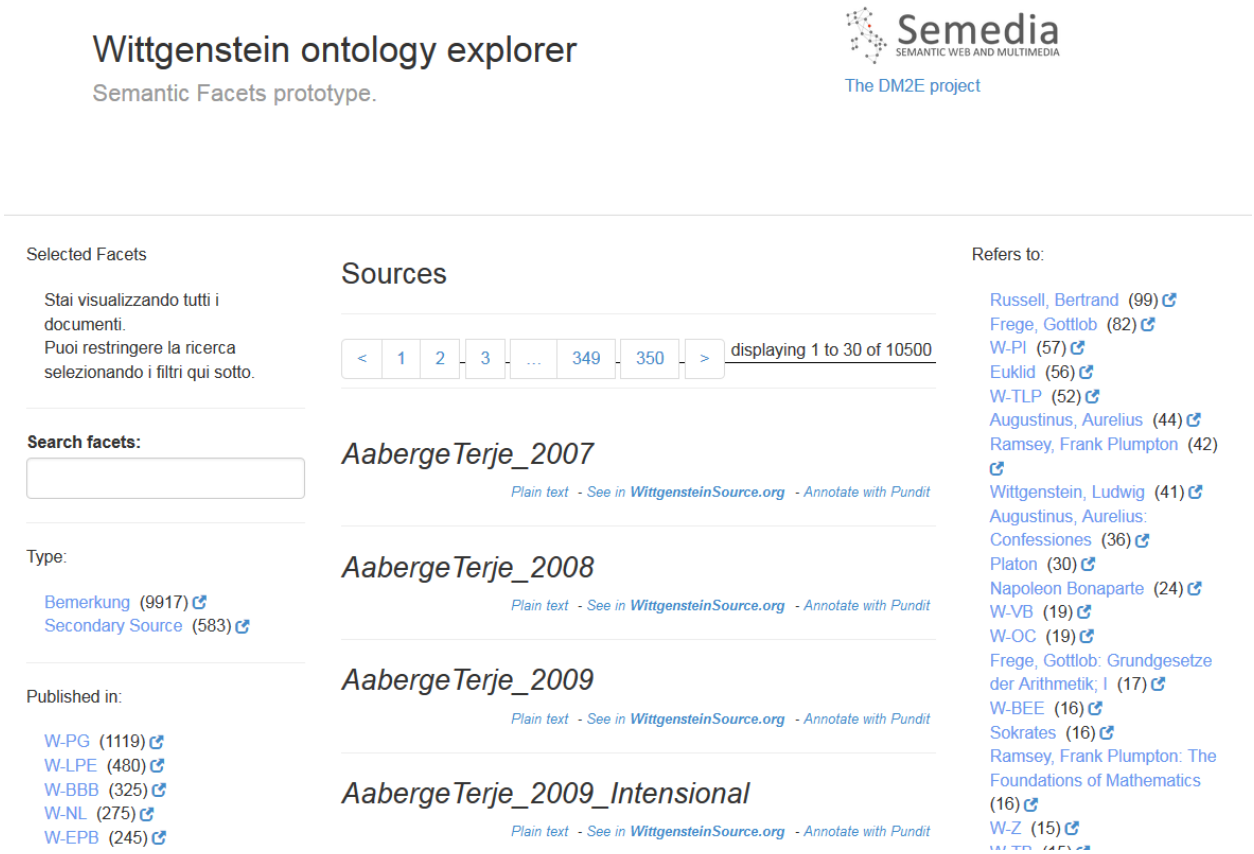

Figure 6: The faceted Wittgenstein Ontology Explorer.

For this use case, we asked two Wittgenstein scholars to use the faceted browser to answer a particular research question about Wittgenstein's oeuvre; they were allowed to choose the question and then given the Self-Documentation worksheet to capture the research/reasoning process. The digitised form of the Nachlass on Wittgenstein Source constituted the basis for both investigations. In following (3.2.2; 3.2.3), the results of both of these experiments documented by the scholars will be described. In addition, the working group carried out open expert interviews with these two scholars about their results. These will also be included (3.2.4), as they were very telling for the purposes of the reasoning experiment.

3.2.2. Scholar One: Wittgenstein's concept of philosophy

The first scholar who completed this experiment (February-April 2014) was already familiar with the ontology, as he is one of the researchers responsible for its creation. His chosen research question and method were partly a simulation of how he imagined another Wittgenstein scholar might use the ontology for its intended purpose. In general, the research question behind the experiment involved imagining how a scholar (perhaps a student) could be assisted by the ontology explorer to come to an understanding of Wittgenstein's conception of philosophy. The associated research method was to explore, analyse and compare primary and secondary sources on the subject by among others exploring key concepts in the ontology. The tools the researcher had at his disposal were the digitised version of Wittgenstein's Nachlass contained on Wittgenstein Source and a faceted browser for exploring the WO, which is linked to the Nachlass.

Finding key texts in Wittgenstein's Nachlass

The first step in the experiment lead the scholar to try to identify key texts relevant for the question what Wittgenstein thinks about philosophy. To do this, he chose the facet "Type" and selected the value "Bemerkung" in the faceted browser.

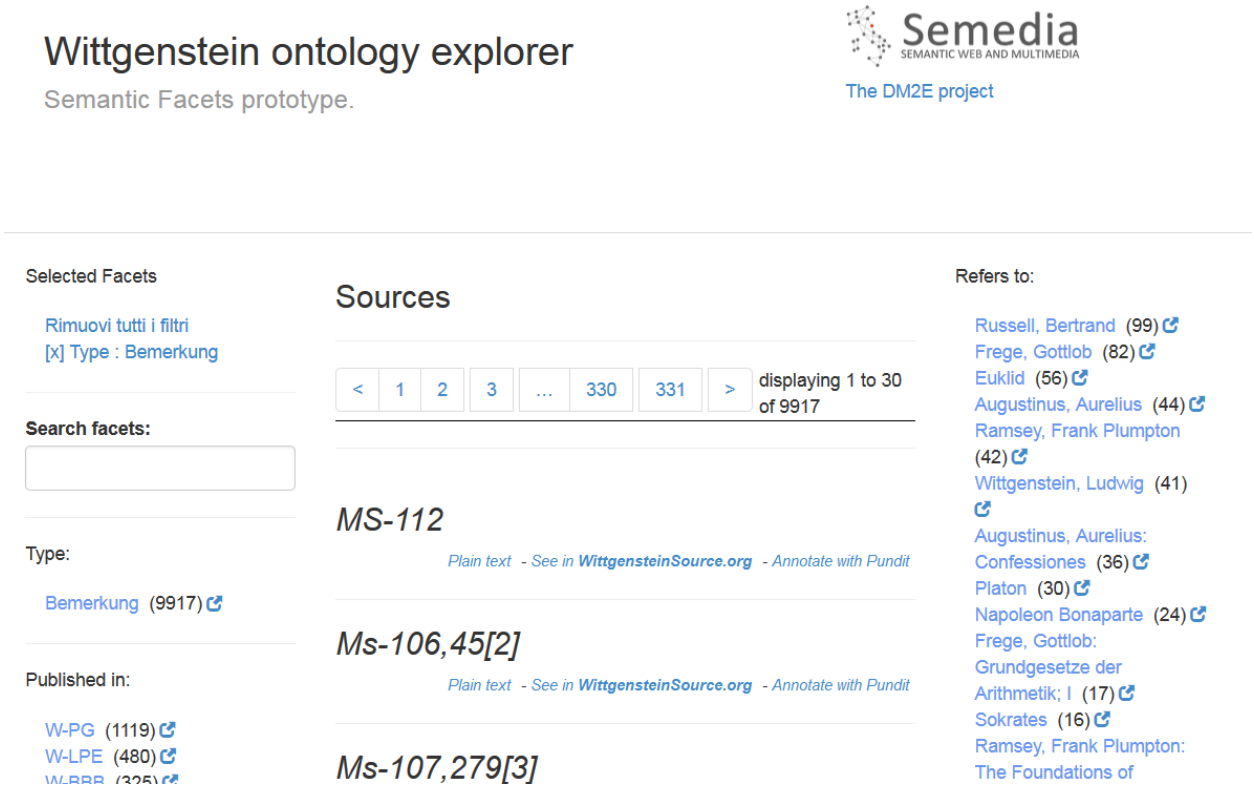

Figure 7: Applying the facet "Type : Bemerkung".

Faceted browsers act as a filter on the data. Choosing a facet value means restricting the search results to only those resources that are associated with that value. By selecting the value "Bemerkung" under the facet "Type", the scholar has eliminated all secondary sources mentioned in the ontology – there are only two "types" in the ontology, see below – from the current view of the faceted browser. "Bemerkung" is an object of the predicate ":hasType". The subject (X) of this triple is the resource representing a text section.

· :X :hasType :Bemerkung

· :X :hasType :Secondary_Source

After this first filtering, the scholar uses the search bar to look for the word "philosophy", and finds out that this word matches one value under the facet "Discusses". In doing so,

the scholar discovers a Bemerkung in Wittgenstein's Nachlass that WAB's Wittgenstein ontology indicates as containing a "discussion" of "philosophy". This passage is in TS-213.

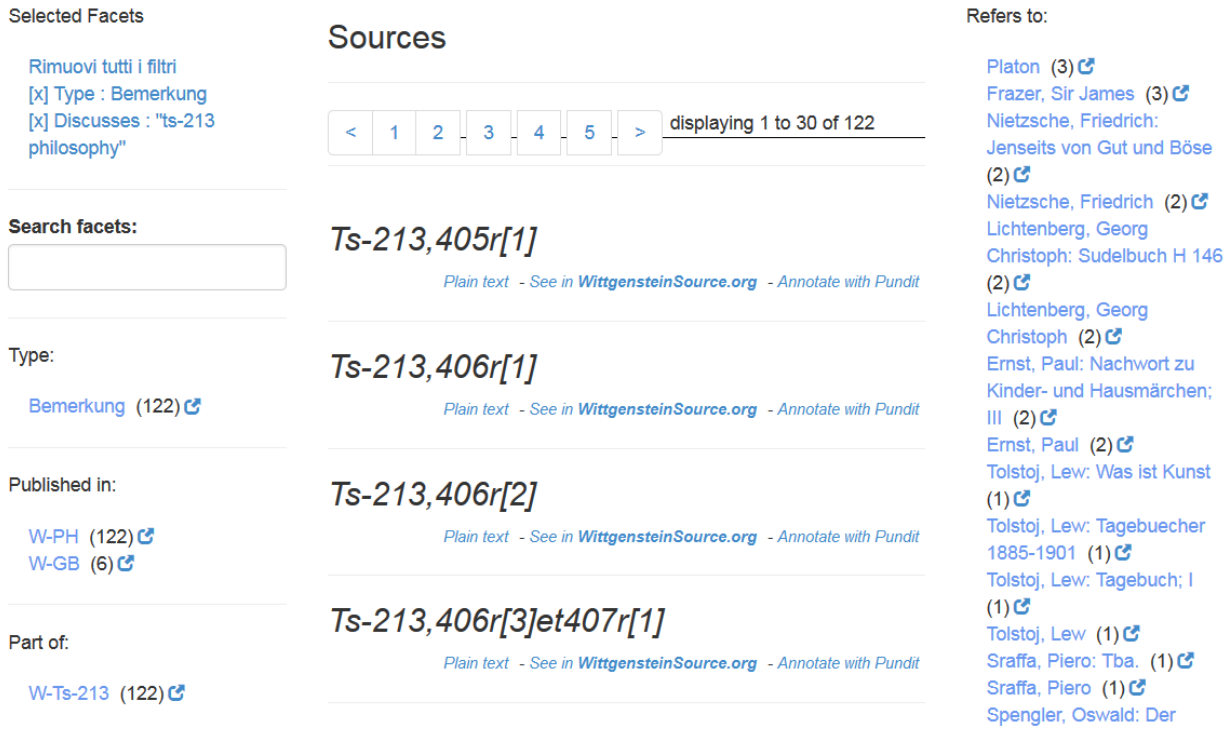

Figure 8: Applying a second facet "Discusses : 'ts-213 philosophy'".

He then selects this facet value, further restricting the search results, and discovers that it brings up several entities (persons, including philosophers) as possible values under the facet "refersTo".

By surveying this list of entities that the graph suggests the Bemerkungen "refers to", the scholar uses his knowledge to come to the conclusion that Wittgenstein's writings were influenced by "continental" rather than "analytical anglo-saxon" conceptions of philosophy. In addition, he assumes that studying / close reading of these sources will give him a clearer idea about the network of concepts Wittgenstein's conception of philosophy is linked to. In other words, he looks at the keywords listed under the "refers

to" facet and assumes that the "refers to" facet would lead him to concepts and philosophers that he would need to study further in order to get a better idea of Wittgenstein's idea of philosophy. Moreover, by adding the facet value "Secondary Source" (which is equivalent to remove the first filter applied) the browser brings up new relevant entities such as articles that "discuss" Bemerkungen themselves containing a "discussion" of philosophy. He can now follow the Web resources linked to the philosophers, and immediately start learning about them as well. With this, the scholar has concluded his short experiment.

There are several elements of this small experiment, which can help us to understand how humanists would like to reason with Linked Data. These include the purpose of the WO itself, the research object and method of the experiment, and the conclusions made by the scholar.

As mentioned before, the creation of ontologies and vocabularies using Linked Data already entails a practice of reasoning, as scholars need them to contain the information and contingencies that will allow for further reasoning with the data. This means that having a look at how and for what purpose vocabularies and ontologies have been created can in and of itself be indicative of the kinds of reasoning scholars want to see enabled by Linked Data. This also means that ontologies and vocabularies can be seen as reflections of research questions and methods in the field and domain for which they have been created. With this in mind, the WO's general purpose, according to its creators is to provide other scholars with tools to assist them in their own research of Wittgenstein's Nachlass, including, among other things, a representation of the key concepts in the corpus and links to secondary sources that may help scholars understand the concepts and ideas expressed in the primary sources. The WO can be considered an attempt to create a knowledge representation or model of the research landscape concerned with Wittgenstein's Nachlass. The ontology appears to reflect on

a research process of close reading and critical analysis of both primary and secondary sources. This process can be aided by the technology, but has to be accomplished by the researcher.

In contrast to the GEI vocabulary, the WO largely captures (explicitly states) only “factual” statements concerning the primary text and excludes, for example, the ontology-creator’s interpretations of the content of the sources. Although choosing key concepts does involve close reading and a certain level of interpretation of the text, this process does not attempt to definitively or explicitly state the meaning of the concept in the text for Wittgenstein research, but merely point to the fact that certain constellations of keywords and sources relevant for scholars are considered to be linkable with certain texts. The creators assume that each scholar will want to partake in close reading and meaning-making using all available sources as well. This indicates that, in the humanities, the meaning to be made from text is variable and dependent on not only the content, but also on the interaction of the individual researcher with the content and with other researchers as well. The WO provides the researcher with a tool to aid this interaction.

The purpose of the ontology is to support the research question explored by the scholar, which was to come to an understanding of Wittgenstein’s conception of philosophy by looking at the key concepts in the text (as catalogued in the WO). The scholar bases this approach among other things on the premise that an overview of the key concepts can help him to gain a quick orientation in the text and to therefore understand the information better. This research question is therefore not only focussed on content, but on a meta-level on the research process as well: He would like to discover if an overview of keywords can tell him something about the content.

The accompanying research process described by the scholar does lead to an answer on both levels. As in the GEI experiment, the scholar is able to obtain new information based on comparing the results list with his real-world knowledge. Although this information is not explicitly contained in the triples, the scholar notices that all of the philosophers "referred to" in the "Bemerkung" in which Wittgenstein discusses his concept of philosophy have something in common: they were so called "continental" philosophers. This is new information for the researcher. He has therefore learned something about Wittgenstein's concept of philosophy. At the same time, the scholar is called to carry out his own close reading, as he might not yet be sure of what these philosophers have said. On the other hand, as the label of these philosophers as "continental" was not contained in the WO and is new information coming *from* the researcher, this information is a valuable addition to the WO and can be recorded in the form of triples via Pundit.

3.2.3 Scholar Two: Wittgenstein's critique of picture theory

The second experiment on the ontology of the WAB was carried out by another scholar who can be considered to be an expert in both Wittgenstein research and ontologies in general. Although not intimately familiar with all of the details of the WO, he is currently working, among other things, on an ontology to represent certain concepts in Wittgenstein's *Tractatus*. On one level, this experiment proved to be unsuccessful, as the scholar was unable to use WO to come to any new conclusions about the data. In fact, he was disappointed with his experience using the ontology browser. This however, led to fruitful discussions with the ontology's creator and the DM2E working group, which will be discussed in 3.2.4. The experiment will therefore be briefly described in this section (3.2.3).

The scholar's initial research question was to use the faceted browser to find what Wittgenstein wrote on the picture theory in the Big Typescript (TS-213). His first step

was to enter the following question in the search bar of the ontology browser in natural language. He entered: “What is Wittgenstein’s critique of the picture theory in Ts-213?”. He then realised that the search bar does not work using natural language.

His second step was then to only enter the expression “picture theory”; he retrieved five results from the secondary sources. In a fourth step he then entered only the expression “picture” and the browser suggested, in his words “many completions or additions”. He then picked the facet "Ts-213-021 Similarity of sentence and picture" and got 12 results. Being a Wittgenstein scholar, he knew that these results were correct and that he could follow them to Wittgenstein Source and read the German text.

His own conclusion was that he could not do very much with the ontology: “This is all I can do with the ontology?”

3.3.4 Results

We invited the second Wittgenstein scholar to discuss his difficulties using the WO explorer with the creator of the WO and documented the resulting discussion. The second scholar was able to uncover certain weaknesses in the implementation of WO on the facets browser as also the underlying dataset. The discussion with the scholars also revealed some basic issues involved in the reasoning process with faceted browsers.

One major issue that was raised in the discussion revolves around trust and authority. As was seen in the GEI experiment, in order to be able to make inferences about the data, humanities scholars need context. This not only includes the context of the dataset itself, but of its creation as well. In this regard, the second Wittgenstein scholar remarked that he would have needed an explanation of the ontology included in the browser in order to be able to understand it fully. The full extent of the underlying dataset was not immediately evident, and the scholar was irritated by the fact that

Wittgentein's *Tractatus* was not included.[42] Seeing as the modelling practice can be related to the scientific method, this includes knowing who made it, for what reason and using what methods and principles. The basis of any further analysis by the scholar relies on the data contained in the ontology being accurate, adequate and, to a certain extent, authoritative.

For this reason, the scholar suggested using a persistent identifier such as the one established in ORCID[43] for each entry in the ontology to identify its creator. Of course, Pundit solves this problem to some extent using the notebook system, but this shows that scholars would like to be aware of this at every step of the process.

One reason mentioned for needing such a persistent identifier of the ontology creator is the division of knowledge in science. It was argued that even Wittgenstein scholars are often experts in only one area such as 'religion and Wittgenstein'. Knowing which scholar is responsible for which information increases transparency; and the knowledge explicated by one scholar for a particular topic will have more weight or authority than another. For an ontology to have stability and authority, it would ideally need area editors.

Another topic discussed in the expert interview revolved around the process of ontology creation itself, which was just as important of a research object for our scholars as Wittgenstein scholasticism. They discussed the need for documentation and standardisation of not only the ontology design, but also of the creation process, so that other scholars or perhaps even machines can understand and recreate it. Both scholars believed that each scholar should be able to question and explore the design of the ontology itself. In this context they saw the ontology browser as a medium of

[42] Due to legal issues.

[43] http://orcid.org/.

communicating one particular view of the domain, which could form the basis of ontological comparison, implicitly as well as explicitly.

4. Conclusion

This paper was primarily set out to inquire “the kinds of reasoning humanities scholars want to see enabled with Linked Data” (DoW). For this purpose, we wanted to foster a perspective on reasoning that is not focused on the aspects computer science is predominantly concerned with, and take into account the prerequisites for the use of Linked Data in the context of interpretative research in the humanities. We therefore proposed the concept “interactive reasoning” as an attempt to approach “reasoning” as a scholarly practice in the context of the Semantic Web.[44] In contrast to automatic inference by machines in the Semantic Web, the term “interactive reasoning” stresses the intention to facilitate reasoning practices for humanists, who conduct their research in the context of Linked Data applications. In this context we specifically concentrated on how humanists can use faceted browsers to explore and reason with Linked Data.

Our method for explaining this issue involved working with three humanities scholars on two particular use cases. In the first use case, we asked a historian to create a small vocabulary for the purpose of semantically annotating a specific corpus of historical textbooks. He then used the PunditSearch faceted browser to query the graph he had created looking for answers to a particular research question. In the second use case, we applied a faceted browser to an existing ontology that was created to be a representation of the research landscape of Wittgenstein’s Nachlass published on Wittgenstein Source. We then had two different Wittgenstein scholars attempt to answer research questions about the Nachlass using the faceted browser. Our first scholar was

[44] Cf. SDM Paper, esp. Interpretative Modeling.

intimately familiar with the WO, our second scholar was an expert in both Wittgenstein and in the topic of ontologies, but was much less familiar with the WO itself.

It was important for us to explore how humanities scholars understand and explore the graph, in particular applications like the Wittgenstein Ontology Explorer to visualise parts of the underlying structured data. And therefore provide a means to the scholar to engage with that structure and subsequently to contribute to it. What Hitchcock (2013) said about the effects of Googling in the field of history also applies to our topic of reasoning and Linked Data: You need to understand what is going on with the graph and how you obtained the results. When scholars create and apply their own data it gives them the necessary context to understand the result.

The experiments[45] were conducted in order to complement the theoretical research on the functional requirements for the translation of Scholarly Operations[46] as well as on the possible application of “reasoning” for scholarship in the humanities in the context of Linked Data with an empirical perspective. The observations uncovered a common threefold structure of the translation and application of interpretative scholarly practices and reasoning to a Linked Data application environment. The first phase includes the initial creation or reuse of vocabulary, to represent the knowledge about certain aspects of the research domain and the methodological approach, to guide and structure the subsequent annotation process. If you understand this first modelling process, you also understand how the researcher plans to reason with the data. The second phase involves the process of interaction with the corpus data, deciding which resources and entities should be annotated and which semantic annotation apply. The third phase involved querying the resulting graph and making inferences about the data that had

[45] Also cf. “Report on Experiments”.

[46] Cf. Scholarly ‘Domain Model.

been created. Therefore, we suggest the following principal and formal three phases of reasoning in Linked Data context:

(1) Conceptualising: Vocabulary selection, modification, or creation which includes the translation of a research interest or question into a Linked Data conceptualisation such as an annotation vocabulary. The vocabulary formalises and explicates the "reasoning" result of a first genuine part of the research process which is based on assumptions or hypotheses about the research to be conducted.

(2) Annotation: The application of the annotation vocabulary to research objects by creating annotations, in the current context with Pundit. Working with the actual texts probably is most commonly associated with the actual research conducted and involves close reading and interpretation, here expressed and formalised as annotation triples.

(3) Exploration: The assessment of the created triples by visualisation, in the current context in a faceted browser which we consider a "low-hanging fruit" for applying one's own reasoning practice to a given knowledge base. Here the researcher explores the previous reasoning and creates new hypotheses which may feed back into the annotation vocabulary and may initiate a new annotation phase.

When analysing our use cases, we noticed that the ontology browser allowed the scholars to quickly gain an insight into certain groups of Linked Data, made by picking certain facets that combined triples of data from a number of sources. The researchers were then able to make inferences about these results, comparing their observations with textual, intertextual and real-world knowledge: the scholar was able to observe, for example, patterns and anomalies implicit in the data. These observations led to the potential creation of new information and therefore new triples. In the GEI use case, our scholar discovered that the historical German textbooks chose to portray the German

Empire as a peaceful country. This was done by not only emphasising the topos of peace, but also by using ‘national’ topics to characterise German. In contrast, other countries were painted as being aggressors. The concept of “national” was new information that could have been added to the corpus to make this new knowledge explicit. In the Wittgenstein Ontology use case, our scholar came to the conclusion that Wittgenstein’s philosophy is influenced by “continental” as opposed to “anglo-saxon” philosophers. This newly inferred information could potentially also be added to the Linked Data vocabulary.

In addition to uncovering one method of coming to conclusions about data using faceted browsers, the experiments also uncovered two very important aspects of reasoning in the digital humanist realm. The GEI use case showed us that context is extremely important. When modelling the domain of German historical school books, GEI attempted to incorporate as many historical facts as possible about the works. This context included not only dates and places, but also facts that might help to uncover biases and values, such as the religious affiliation of the authors. This played a large role in his vocabulary, which tried to make certain value concepts explicit. In the WO use case, the conversation between our scholars revealed that the reasoning process relies heavily on the context of the research itself. This means scholars need to know who is creating what annotations for what purposes and how. As a result of our use cases, we have determined that context on all levels needs to be taken into consideration when considering reasoning scenarios with Linked Data in the digital humanities.

In conclusion, the three step process discussed here represents the basic structure of one way of how humanists do and want to reason in the context of Linked Data and interpretative approaches. The same basic iterative 3-tiered process has been identified in other experiments we conducted. Of course, Pundit predetermines the outcomes of the

experiments to a certain extent in terms of the available functionality. However, we nevertheless found this three step procedure as fruitful to represent simple interpretative approaches from the humanities in a Linked Data context. All experiments demonstrated its potential. In that regard, we propose further systematic research be conducted based on this principle approach in order to deepen our understanding of how humanists do and want to reason, and more generally conduct parts of their research, in the Linked Data and Semantic Web context.

Acknowledgments

We would like to thank the following persons for their support in the experiments and their contributions to this paper: Alessio Piccoli (Net7), Andreas Weiß (GEI), Jakub Mácha (Masaryk University, Czech Republic)

Notes

References

Beynon, M.; Russ, S. and McCarty, W. (2006). Human Computing: Modelling with Meaning, Literary and Linguistic Computing, 21 (2), pp. 141–157: 10.1093/llc/fql015 (accessed 31 January 2015).

Bowen, W. R. and Siemens, R. G. (eds) (2008). New technologies and Renaissance studies: Contributions from annual conferences held 2000 to 2005 during gatherings of the Renaissance Society of America. Tempe, AZ: Iter.

Blanke, T. and Hedges, M. (2013). Scholarly Primitives Building institutional infrastructure for humanities e-Science, Future Generation Computer Systems 29 (2): 654–661: 10.1016/j.future.2011.06.006 (accessed 31 January 2015).

Crombie, A. C. (1994). Styles of scientific thinking in the European tradition: The history of argument and explanation especially in the mathematical and biomedical sciences and arts. London: Duckworth.

Deegan, M. and Sutherland, K. (2009). Transferred illusions: Digital technology and the forms of print. Farnham: Ashgate.

Dentler, K. et al. (2011). Comparison of Reasoners for large Ontologies in the OWL 2 EL Profile. In Semantic Web 1 (1-5). http://www.semantic-web-journal.net/sites/default/files/swj120_2.pdf (accessed 31 January 2015).

Eiter, T. and Krennwallner, T. (eds) (2012). Reasoning Web. Semantic Technologies for Advanced Query Answering: Proceedings of the 8th International Summer School, Vienna, Austria, September 2012. Berlin: Springer. (Lecture Notes in Computer Science, 7487).

Gardiner

Gradmann, S. (2013). Semantic Web und Linked Open Data. In Kuhlen, R., Semar, W. and Strauch, D. (eds), Grundlagen der praktischen Information und Dokumentation. Handbuch zur Einführung in die Informationswissenschaft und -praxis. Berlin: De Gruyter Saur, pp. 219–228.

Grassi, M. et al. (2012). Pundit: Semantically Structured Annotations for Web Contents and Digital Libraries. In Mitschick, A. et al. (eds), Semantic Digital Archives 2012: Proceedings of the Second International Workshop on Semantic Digital Archives,

Paphos, CY, September 2012. http://ceur-ws.org/Vol-912/paper4.pdf (accessed 31 January 2015).

Grassi, M. et al. (2013). Pundit: augmenting web contents with semantics. Literary and Linguistic Computing 28 (4): 640–659: DOI: 10.1093/llc/fqt060 (accessed 31 January 2015).

Gruber, T. R. (1993). Toward Principles for the Design of Ontologies Used for Knowledge Sharing, International Journal of Human-Computer Studies 43 (4-5), pp. 907–928.

Hacking, I. (1985). Styles of Scientific Reasoning. In Rajchman, J. and West, C. (eds), Post-analytic philosophy. New York: Columbia University Press, pp. 145–164.

Hitchcock, T. (2013). Confronting the Digital, Or How Academic History Writing London the Plot. In Cultural and Social History 10 (1): 9-23: 10.2752/147800413X13515292098070 (accessed 31 January 2015).

Holyoak, K. J. and Morrison, R. G. (eds) (2012). The Oxford Handbook of Thinking and Reasoning. Oxford: Oxford University Press.

Kuhlen, R.; Semar, W. and Strauch, D. (eds) (2013). Grundlagen der praktischen Information und Dokumentation: Handbuch zur Einführung in die Informationswissenschaft und -praxis. Berlin: De Gruyter Saur.

Ludwig, S. A. (2010). Comparison of a Deductive Database with a Semantic Web Reasoning Engine. Knowledge-Based Systems 23 (6): 634–642: 10.1016/j.knosys.2010.04.005 (accessed 31 January 2015).

McCarty, W. (2005). Humanities computing. Houndmills: Palgrave Macmillan.

McCarty, W. (2012). The residue of uniqueness, Historical Social Research 37 (3): 24–45. http://www.jstor.org/stable/i40077811 (accessed 31 January 2015).

McCarty, W. (2008). Being Reborn: The Humanities, Computing and Styles of Scientific Reasoning. In Bowen, W. R. and Siemens, R. G. (eds), New technologies and Renaissance studies. Tempe, AZ: Iter. (New technologies in medieval and renaissance studies, 1), pp. 1-22.

Meister, J. C. (ed) (2012). Digital Humanities 2012: Conference Abstracts, University of Hamburg, July 16-22. Hamburg: University Press.

Oldman, D., Doerr, M. and Gradmann, S. (n.d.). ZEN and the Art of Linked Data. New Strategies for a Semantic Web of Humanist Knowledge. To be published in Schreibman, S., Siemens, R. G and Unsworth, J. (eds), A new Companion to Digital Humanities. Oxford: Blackwell [preprint].

Pesce, M. (1999). SCOPE1: Information vs. Meaning. http://hyperreal.org/~mpesce/SCOPE1.html (accessed 31 January 2015).

Pichler, A. and Zöllner-Weber, A. (2012). Sharing and debating Wittgenstein by using an ontology [doi: 10.1093/llc/fqt049]. In: Literary and Linguistic Computing, Vol. 28 (2013) / Number 4. pp. 700-707. (UK) Oxford: Oxford University Press.

Rajchman, J. and West, C. (eds) (1985): Post-analytic philosophy. New York, NY: Columbia University Press.

Svensson, P. (2009): Humanities Computing as Digital Humanities, In Digital Humanities Quarterly 3 (3). http://www.digitalhumanities.org/dhq/vol/3/3/000065/000065.html (accessed 31 January 2015).

Zoglauer, T. (2008): Einführung in die formale Logik für Philosophen. Göttingen: Vandenhoeck & Ruprecht.

Zöllner-Weber, A. (2009): Ontologies and Logic Reasoning as Tools in Humanities?, In Digital Humanities Quarterly 3 (4). http://www.digitalhumanities.org/dhq/vol/3/4/000068/000068.html (accessed 31 January 2015).

ADDENDUM

Addendum 1: Guidelines

Guidelines for the documentation of DH reasoning practices in the context of DM2E and the tools Pundit and the faceted browser in Ask

The principle goal of this experiment is to observe the information behaviour of humanists while working with Linked Data. The technical and conceptual scope of the experiment is a faceted browser which allows to explore triple data. The particular research interests focuses on the work and reasoning process the humanist applies within this particular setting while trying to find answers to particular research questions relevant to their domain of discourse.

The outcomes of this experiment will provide empirical evidence for the type of reasoning humanists want to apply to triple data.

*The participants are asked to choose **two** research questions which are **relevant** for their particular domain of discourse and which they expect to be **applicable** to the corpus at hand!*
The participants will then try to answer these research questions using the faceted browser and create a self-documentation of the technical and reasoning procedure they applied.

The documentation should address the following four sections: The first section provides a brief overview on the characteristics of the corpus (i.e. the data) you are using for the experiment. The second section gives a brief description for each research question you will apply to the corpus. The third section provides a guideline for recording your work process for each research question in the faceted browser. The fourth section summarizes your experiences and considers proper reasoning scenarios. by answer the following questions:

1. Corpus

Please provide a short overall description of the corpus and its Linked Data representation you are using!

- Who created the corpus, for whom and for what purpose?
- What kinds of data does it contain (eg. annotations, ontology, textual data, digitized images etc.)?
- Who created the linked data representation of the corpus, for whom and for what purpose?

- Describe the faceted browser! What kinds of entities and relations did you markup in Linked Data (what vocabularies and ontologies did you use) to create your facets?
- *How many triples were incorporated in the faceted browser*

2. Research questions

For each research question please provide a brief general characterisation!

- Why did you choose this particular research question? What is the relevance of the research question for the particular domain of discourse your corpus is addressing?
- Which answers do you expect and why do you think the corpus will provide sufficient information?

3. Protocol

For each research question create a step-by-step protocol ("Verlaufsprotokoll") of each step you are taking during your work with the faceted browser. This protocol should also support you in writing up a summary of your process and to assess and justify the steps you took. The main point is that you try to be self-conscious about which kinds of assumptions and conclusions you are drawing along the way.

For each step, try to describe in a detailed manner each single action/step you took and then try to explain why you performed the action/step (consider the guiding questions!).

Describe the process!	**Reflect on the Process!**
Research question / problem:	
Step/Action 1:	

Step/Action 2:	
Step/Action 3:	
Step/Action 4:	
Step/Action 5: As	
Step/Action 6:	
Step/Action 7:	

4. Summary

(a) For each research question please provide a summary of your process by considering the following questions.

- What is the final answer you found for the research question?
- Assess the quality and usefulness of the answer you found from a scientific viewpoint!
- Compare (a) the reasoning process, or aspects of this process, you applied during your work with the faceted browser to (b) the reasoning process you would apply when working in an mostly non-digital setting.

(b) For each research question, please provide a reflection on the reasoning process involved.

- Did the premises you had before starting working on the research question influence the way you reasoned or proceeded? If so how?
- How would you describe the research method/process used in answering the research question?
- How did the assumptions you made about the data and the research methods inform the conclusions you came to?
- Do you presume the results to be trustworthy or do you have doubt about their trustworthiness?

(c) As an overall consideration please provide your viewpoint on the potential of proper reasoning and inference scenarios for your particular use case.

- Discuss the potential of reasoning software for answering the particular research question!
- Did any new questions arise from the results generated?
- Which conceptual or technical aspects you encountered while using the faceted browser influenced your work and how?